

Citation for published version:

Petrini, K, McAleer, P, Neary, C, Gillard, J & Pollick, FE 2014, 'Experience in judging intent to harm modulates parahippocampal activity: An fMRI study with experienced CCTV operators', *Cortex*, vol. 57, pp. 74-91.
<https://doi.org/10.1016/j.cortex.2014.02.026>

DOI:

[10.1016/j.cortex.2014.02.026](https://doi.org/10.1016/j.cortex.2014.02.026)

Publication date:

2014

Document Version

Peer reviewed version

[Link to publication](#)

University of Bath

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Experience in judging intent to harm modulates parahippocampal activity: An fMRI study with experienced CCTV operators

Karin Petrini¹, Phil McAleer², Catherine Neary³, Julia Gillard², Frank E Pollick²

¹Institute of Ophthalmology, University College London, London, United Kingdom

²School of Psychology, University of Glasgow, Glasgow, United Kingdom

³Advanced Technology Centre, BAE Systems, Bristol, United Kingdom

Corresponding Author:

Dr Karin Petrini

Postdoc Research Associate

UCL Institute of Ophthalmology

11-43 Bath Street

London EC1V 9EL

United Kingdom

tel: +44 (0)20 76084015

Email: k.petrini@ucl.ac.uk

Running Title

CCTV experience modulates parahippocampal activity

Abstract

Does visual experience in judging intent to harm change our brain responses? And if it does, what are the mechanisms affected? We addressed these questions by studying the abilities of CCTV operators, who must identify the presence of hostile intentions using only visual cues in complex scenes. We used functional magnetic resonance imaging to assess which brain processes are modulated by CCTV experience. To this end we scanned 15 CCTV operators and 15 age and gender matched novices while they watched CCTV videos of 16 seconds, and asked them to report whether each clip would end in violence or not. We carried out four separate whole brain analyses including 3 model-based analyses and one analysis of intersubject correlation to examine differences between the two groups. The three model analyses were based on 1) experimentally predefined clip activity labels of fight, confrontation, playful, and neutral behaviour, 2) participants' reports of violent outcomes during the scan, and 3) visual saliency within each clip, as pre-assessed using eye-tracking. The analyses identified greater activation in the right superior frontal gyrus for operators than novices when viewing playful behaviour, and reduced activity for operators in comparison with novices in the occipital and temporal regions, irrespective of the type of clips viewed. However, in the parahippocampal gyrus, all three model-based analyses consistently showed reduced activity for experienced CCTV operators. Activity in the anterior part of the parahippocampal gyrus (uncus) was found to increase with years of CCTV experience. The intersubject correlation analysis revealed a further effect of experience, with CCTV operators showing correlated activity in fewer brain regions (superior and middle temporal gyrus, inferior parietal lobule and the ventral striatum) than novices. Our results indicate that long visual experience in action observation, aimed to predict harmful behaviour, modulates brain mechanisms of intent recognition.

Keywords: CCTV experience, parahippocampal cortex, predicting violence, fMRI, eye-tracking.

Introduction

The use of Closed Circuit Television (CCTV) to monitor human activity has become a ubiquitous feature in societies worldwide. These systems typically employ a set of cameras deployed around a complex urban geography, and these are monitored and controlled from a central location. One potential benefit of CCTV is to identify when hostile intentions are present, to allow the formation of a pre-emptive response that minimizes the consequences of these intentions (Diffley and Wallace, 1998; Weinberger, 2010; Welsh and Farrington, 2002). However, to be effective, this application of CCTV requires a human operator to be able to identify the presence of hostile intentions from only visual cues in complex scenes (Howard et al., 2011a; Howard et al., 2011b; Howard et al., 2009; Stedmon et al., 2008; Troscianko et al., 2004). The amount of visual information and the number of actions these operators monitor every day is very high, and this makes these individuals optimal subjects for examining how extensive experience in the visual analysis of behavior in real-world scenes affects human brain function.

Quickly understanding the intentions of others based solely on visual cues relies on different abilities, which are supported by different brain regions and mechanisms. For example, CCTV operators may need to adopt a rapid simulation routine. It is suggested that a mirror neuron system (e.g. Decety and Grezes, 1999; Gallese et al., 1996; Nishitani and Hari, 2000; Rizzolatti et al., 1996) or an “action observation network” (e.g. Grafton, 2009) may exist in the human brain, underlying imitation and simulation capabilities, and activating both when an agent performs an action and when they observe the same action performed by another agent. To date a number of studies have described the participation of the premotor cortex, parietal, inferior frontal and STS areas when observing and/or producing human actions; findings obtained by using functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and transcranial magnetic stimulation (TMS) techniques (Buccino et al., 2001; Fadiga et al., 2005; Fadiga et al., 1995; Grafton et al., 1996a; Grafton et al., 1996b; Iacoboni et al., 2001; Keysers and Gazzola, 2006; Rizzolatti et al., 1996; Strafella and Paus, 2000). Others have described a somatotopic organization in these areas when either observing or producing movements of different body parts (Buccino et al., 2001). Further, familiarity and expertise with the observed and/or performed actions has been shown to modulate activity in these action recognition areas (e.g. Calvo-Merino et al., 2005; Calvo-Merino et al., 2006; Cross et al., 2006; Cross et al., 2009; Petrini et al., 2011).

Quick detection of harmful intentions may also require the attribution of mental states to others. This ability has been repeatedly attributed to a different group of brain areas, which

are described as underlying the capacity to understand the thoughts and beliefs of others, as though we could read their minds (Amodio and Frith, 2006; Van Overwalle and Baetens, 2009). This mentalizing system, also known as theory of mind (ToM), includes the precuneus (PC), the temporoparietal junction (TPJ) and the medial prefrontal cortex (mPFC, Amodio and Frith, 2006; Frith and Frith, 1999; Mitchell, 2006; Saxe, 2006). It has been examined in healthy novices as well as patient populations, using fMRI and PET methods (Castelli et al., 2002; Frith and Frith, 2006; Pelphrey et al., 2004; Pelphrey et al., 2003; Saxe and Kanwisher, 2003; Saxe et al., 2004; Schultz et al., 2003). For example, fMRI studies examining biological motion have reported diminished activity in the right TPJ in people with ASD (Herrington et al., 2007; Kaiser et al., 2010).

CCTV operators need to gather relevant visual information while retrieving episodic memories (i.e. memory of autobiographical events). The role of the hippocampus in retrieving episodic memories and processing spatial relationships, and that of the parahippocampal areas in processing viewpoint-specific scene representations and spatial navigation, have now been confirmed by several blocked (Epstein et al., 1999; Stern et al., 1996) and event-related (Kirchhoff et al., 2000; Soto et al., 2007; Turk-Browne et al., 2006) fMRI studies. Some of these studies have demonstrated reduced response (i.e., adaptation) in the parahippocampal cortex when comparing activity elicited by repeated visual scenes with that elicited by novel scenes. Consistent with this are claims that the parahippocampal cortices are activated in working memory tasks involving novel but not familiar stimuli (Hasselmo and Stern, 2006). Additionally, the parahippocampal gyrus has been found to show specific activation for action prediction as opposed to memorizing action information (Stadler et al., 2011), indicating that this area might have a bigger role in action recognition than memory storage. The fact that the parahippocampal gyrus has been found to change its activity in response to familiar and frequently executed actions (e.g. Calvo-Merino et al., 2005; Petrini et al., 2011) and to have neurons that respond to action observation, execution or both, led to the suggestion that it may be part of an extended “mirror neuron” system (Mukamel et al., 2010).

Being exposed to violent and aggressive media has been shown to reduce activation in frontal and parietal regions of the brain, suggesting that repeated exposure to aggressive videos causes emotional desensitisation in adults and adolescents (Mathiak and Weber, 2006; Strenziok et al., 2011). For example, Mathiak and Weber (2006) showed that experienced male gamers recruited the orbital frontal cortex less during a violent first-person shooter game than during a non-violent game. Further regions showing deactivation during virtual

violence included the right anterior cingulate cortex, the amygdala, the parahippocampal area and the intraparietal sulcus. CCTV operators, who are repeatedly exposed to violent events, may employ a similar adaptive mechanism by suppressing emotional responses while they operate the surveillance cameras to obtain evidence.

All of these different networks may be involved during CCTV operators' surveillance, and examining whether and how experience in judging others' intentions influences these networks can help us characterize their different relevance and functions during such complex everyday tasks. In the present study, we used functional magnetic resonance imaging (fMRI), eye movement and behavioral methods to investigate the effects of experience on viewing CCTV footage. Our main hypothesis was that prolonged sensory experience would modulate neural representations of others' intents in either one or all of the aforementioned mechanisms; more specifically, we hypothesized a reduction in brain activity for CCTV operators when compared to novice viewers, in line with recent studies showing the effect of experience on brain deactivation (Berkowitz and Ansari, 2010; Brewer et al., 2011; Kelly and Garavan, 2005; Petrini et al., 2011).

Using a series of behavioral (e.g. rating of violent outcome) and eye-tracking measures, we selected 24 short videos of 16 seconds long to use as stimuli during the fMRI experiment. The experience of watching videos of complex activity and scenes was very similar to the CCTV operators' real-life jobs, making this group very appropriate for the study of visual experience in real-world situations. During fMRI scanning, 15 CCTV operators and 15 novice viewers watched the clips and judged whether they thought the activity would or would not result in violence. None of the operators were familiar with the particular urban environments from which the clips were obtained. The data were analyzed in three ways using model-based analysis. In the first analysis, by Clip Type, we used the pre-defined labels of 'fight', 'confront', 'play' and 'nothing' to distinguish between the different conditions and to allow us to explore the effect of this variable on novice and operator brain responses. In the second analysis, we ignored these pre-defined experimenter labels and used the responses given by the participants, in order to explore the differences between the novices and operators when they did and did not predict violence. In the third analysis, we compared novice and operator brain activity at the time point that a preliminary eye movement experiment had defined as being the most salient. Finally, we performed an intersubject correlation analysis (Hasson et al., 2004) to assess the effect of CCTV experience on CCTV operators' neural timecourses as they watch the same natural clips. To achieve this we

performed a whole brain analysis comparing operator-to-operator intersubject correlation with novice-to-novice intersubject correlation.

Methods

Participants

Behavioral phase before scanning and stimulus selection

In total, 21 participants (11 CCTV operators, 3 female, 8 male, aged 21-53 years ($M = 36.3$, $SD = 10.1$); and 10 novices, 2 female, 8 male, aged 28-43 years ($M = 33.8$, $SD = 6.0$)) were recruited to take part in the experiment. No significant difference was found between the age of operators and novices (independent t -test, $t(19) = 0.875$, $p = 0.392$). The ‘operator’ participants were defined as individuals with CCTV surveillance experience and they were recruited from CCTV control rooms and user groups. The operator participants were paid for their time and travel expenses to attend the experiment at BAE Systems, Advanced Technology Centre, Bristol. Each participant read and signed a Consent Form that described their participation in the experiment and the use of the data collected. All participants were free to leave the study at any time. The ‘novice’ participants were defined as individuals with no CCTV surveillance or security experience. The overall mean experience of CCTV monitoring among the CCTV operators was 4.5 years (Standard Deviation (S.D.) = 3.0, range 0.4-12 years) and on an average day, the overall mean time for which an operator viewed CCTV footage was 10 hours (range 8-12 hours). All the operators were currently working, and the years of experience reported here refer to those prior to the date of scan. Ethical approval for all phases of the study was obtained from the UK Ministry of Defence Research Ethics Committee.

fMRI phase

A new and different group of 30 healthy individuals participated in the fMRI phase of the experiment: 15 CCTV operators (age range 38-57 years; $M = 47.13$; $S.D. = 7.01$, 4 women, 10 right-handed) with a minimum of 6 months experience (range 0.8-17 years; $M = 10.35$; $S.D. = 5.33$), and 15 individuals with no experience of CCTV surveillance in their workplace (age range 38-59 years; $M = 47.73$; $S.D. = 6.32$, 4 women, 13 right-handed). The higher number of male than female operators reflects a real difference in number of male and female employed as CCTV operators (see also Troscianko et al., 2004). Consequently, we recruited the same number of age and IQ matched female novices. No significant difference was found between the age of operators and novices (independent t -test, $t(28) = 0.246$, $p = 0.807$). All

the operators were currently working, and the years of experience reported here refer to those prior to the date of scan. Participants provided informed consent to take part in the study, which was approved by the UK Ministry of Defence Research Ethics Committee. Gender was matched for the CCTV operators and novices who took part in this fMRI study. After scanning, participants completed an Aggression Questionnaire, an Interpersonal Reactivity Index (IRI) and two word processing assessments to measure IQ: Spot-The-Word (measuring ability to recognize true words and non-words) and Speed of Comprehension (measuring speed of reading and understanding sentences in two minutes). The Aggression Questionnaire (Buss and Perry, 1992) is composed of four positively correlated subscales (physical aggression, verbal aggression, anger, and hostility) where participants indicate to what extent statements reflect their character on a 5-point scale. Results showed no significant differences between operators and novices on the Aggression Questionnaire (physical aggression: $t(28) = 0.304, p = 0.763$ (operators: $M = 18.33$ and $S.D. = 5.61$; novices: $M = 19$ and $S.D. = 6.36$); verbal aggression: $t(28) = -1.132, p = 0.267$, operators: $M = 14.6$ and $S.D. = 2.84$; novices: $M = 13.33$ and $S.D. = 3.27$; anger: $t(28) = 0.840, p = 0.408$, operators: $M = 12.87$ and $S.D. = 5.54$; novices: $M = 14.4$ and $S.D. = 4.39$; hostility: $t(28) = -1.124, p = 0.271$, operators: $M = 17.8$ and $S.D. = 4.44$; novices: $M = 16.07$ and $S.D. = 3.99$). The IRI (Davis, 1980) is a commonly used empathy measure composed of questions on four subscales (Perspective Taking, Fantasy, Empathetic Concern, and Personal Distress) where participants indicate to what extent each statement reflects their character on a 4-point scale. No significant differences were found between CCTV operators and novices on three sub-scales of the IRI (fantasy scale: $t(28) = -0.987, p = 0.332$, operators: $M = 13.4$ and $S.D. = 3.31$; novices: $M = 11.8$ and $S.D. = 5.33$; empathetic concern: $t(28) = 0.731, p = 0.471$, operators: $M = 18.73$ and $S.D. = 5.14$; novices: $M = 20$ and $S.D. = 4.31$; personal distress: $t(28) = 1.011, p = 0.321$, operators: $M = 9.27$ and $S.D. = 3.99$; novices: $M = 10.8$ and $S.D. = 4.31$), but a significant difference was found for perspective taking ($t(28) = 2.121, p = 0.043$, operators: $M = 16$ and $S.D. = 5$; novices: $M = 19.4$ and $S.D. = 3.68$), with the novices having higher scores. The Spot-the-Word Intelligence test (Baddeley et al., 1993) presents subjects with 60 pairs of items, where each pair is made up of a word and a non-word, and participants are instructed to indicate which one of the pairing is a real word. This short IQ test has been shown to give a good indication of a participant's intelligence in comparison with the general population. Results showed no significant difference in scores ($t(28) = 1.143, p = 0.263$) between CCTV operators ($M = 47.7, S.D. = 4.7$) and novices ($M = 50.1, S.D. = 6.3$) on the Spot-The-Word test, with both groups performing on average at approximately the 50th percentile of

intelligence. Mean scores on the Speed of Comprehension test were also very similar ($t(28) = 0.178, p = 0.860$), with no significant difference being found between the operators ($M = 63.1, S.D. = 18.1$) and novices ($M = 64.3, S.D. = 16.8$). CCTV operators were all shift workers, while the novices were not all shift workers. To minimise effects that might arise simply from attention alone, we intentionally used an experimental task with low attentional demands (i.e. we used short clips, only a single screen). We also considered using a matched profession that included shift work, but we were unable to come up with one that appeared a perfect match. In the end we choose a heterogeneous mixture of professions for the novice group (e.g. house Painter, chef, brain injury project co-ordinator, fireman, taxi driver, managing director) that was age and IQ matched.

We also used linear regression analysis to examine whether the rating of perspective taking and empathic concern changed with the age of the participants or the operators' years of CCTV experience. We found only one significant inverse correlation: between CCTV operators' years of experience and empathic concern ($r = -0.499, p = 0.029$, one-tailed; $r = -0.499, p = 0.058$, two-tailed). Hence, CCTV operators gave lower ratings of perspective taking than novices, and CCTV operators with greater experience gave even lower ratings of empathic concern than CCTV operators with less experience.

Stimuli

Originally, over 800 hours of CCTV footage was obtained of urban scenes in the UK. For the selection process, four viewers were asked to identify incidents that could be classified as aggressive behavior and actual fights. These viewers were all paid research assistants with no prior CCTV experience who screened the corpus of video material under the supervision of two of the authors (PM & FEP). Sixty incidents of varying durations were found, and it was decided that only CCTV clips which had at least 16 seconds of aggressive behavior leading up to a violent incident for the fight clips would be used, as the aim of the experiment was to investigate the prediction, not recognition, of violence. After those incidents with less than 16 seconds of build-up were discarded, Adobe Premier Pro 1.5 was used to alter the remaining scenes to show only the 16 seconds prior to physical contact. Those CCTV clips that resulted in physical aggression (and therefore included hostile intent) were labelled as the 'Fight' clips. After selecting the Fight clips, control scenes were chosen for the 'Confront', 'Play', and 'Nothing' categories. The CCTV clips obtained for the other three categories were matched to the Fight clips in several respects: location, time of day, and number of people in each display (see Troscianko et al., 2004 for similar matching criteria). If a match could not be obtained to a potential Fight clip then it was not used in the study. The main differences

between the clip categories related to whether a violent incident occurred after the clip ended (only in the Fight clips), whether aggressive behavior was shown throughout the clip (Fight and Confront clips) and whether physical interactive behavior was shown in the clips (Fight, Confront, and Play clips). Aside from the four employed viewers, no participant involved in this study viewed any actual violent acts during these 16-second displays. A pilot behavioral experiment was performed to assess the ability of novice and experienced observers to detect hostile intent from these displays. This resulted in the selection of 9 displays for each of the four different categories, making a total of 36 displays. All 36 displays were used for the eye movement and behavioral measurements that preceded the fMRI study. However, for the fMRI study the number of displays was reduced to 6 for each condition, by selecting the displays that were most reliably discriminated between by both novices and operators. Due to privacy issues we cannot show the stimuli. This because the footage we used in our studies contained sensitive personal information that cannot be shared in a public forum. This requirement was given to us by the bodies which provided us with the footage and agreed to by the ethics panel approving our research.

Design and procedure

Behavioral phase before scanning

Nine 16-second CCTV clips from each of the four categories (Fight, Confront, Play, and Nothing) were used. Each participant was shown the clips in a quasi-random block design. Two opposite orders of clips presentation were used for run1 and run2 in the pre-scan phase as well as during the scan. These orders were kept constant across subjects although were counterbalanced in terms of which run was shown first (i.e. one operator saw run1 then run2 and another operator saw run2 then run1 etc....and the same for novices). It was decided that the genetic algorithm (Wager and Nichols, 2003) was not an effective solution to create these runs as it is designed for rapid event-related runs rather than the block type designs similar to the one we use here. So we opted for pseudo-randomised runs that minimised any preceding bias. That is, no display was ever preceded by another clip from the same category (e.g. a fight clip was never preceded by another fight clip) or by that of another category more than once. Participants' eye movements were recorded while they watched each of the 36 CCTV clips. Fixations, saccades and scan paths are three useful measures of visual behavior and eye movements. Although the mean fixation durations, gazing times, and saccade rates were calculated, we were primarily interested in the time and sequential orders of fixation-based measurements, because they would show how individuals gathered information during the scene. We additionally constructed a measure of the variability of eye movements to find the time and spatial location when the observers' eye

gazes were most in agreement with each other. This was defined as the time and location of a group of scan paths where the least variability of position was obtained. This measure was termed the 'salience measure', since we assumed that when the observers were all looking at the same visual point at the same time, this would correspond to a salient event in the clip (e.g. Sawahata et al., 2008; Smith and Mital, 2013; Stainer et al., 2013). After each clip was shown during the eye-tracking experiment, the participant was asked to rate the likelihood that a violent incident occurred after it ended, using a six-point scale that ranged from 1 (extremely unlikely) to 6 (extremely likely). Once the eye-tracking part of the experiment was complete, the participants viewed all the clips a second time and immediately after each clip they were asked to rate the perceived level of hostility, using a ten-point rating scale. This scale ranged from 1 (very jovial, no hostility), through 5/6 (neutral, benign, no emotions present), to 10 (very hostile, aggressive).

fMRI phase

Every individual in the experiment underwent the same procedure. After providing informed consent they were taken to the scanner and underwent two runs of scan in one session. During this time, each participant 1) viewed the clips while being scanned and provided a behavioral response which demonstrated whether or not they believed the clip would end in violence at the end of each 16-second display; and 2) provided any general feedback on their experience during the experiment. Eye movements were monitored by a ViewPoint Eye Tracker, PC-60 Software (Arrington Research) to ensure participants were paying attention and to determine the direction of gaze. Unfortunately, however, technical difficulties limited the amount of gaze data recorded and this precluded its use as part of the general analysis.

fMRI experiments: stimulus delivery

A Windows PC coordinated the display of the experimental video materials, with stimulus presentation and response collection being controlled using the software package Presentation, Neurobehavioral Systems (www.neurobs.com). Each display was a 16 second video that had been converted into an AVI format movie with a resolution of 480×576 pixels. The movies were presented to participants in the center of the screen, using Nordic Neurolabs Visual System goggles with a field of view of $30^\circ \times 22.5^\circ$. The movies filled a visual angle of approximately $19^\circ \times 22.5^\circ$. The 24 trials were randomly distributed between two runs of 12 trials each. Each run began with 16 seconds of blank screen followed by 10 seconds of fixation towards a central target before the video trials began. These sections of

fixation were inserted as a means of examining calibration of the eye tracker across the block, however due to other issues with the eyetracking this feature could not be exploited. Each trial began with the 16-second video depicting one of the four conditions, and participants responded with a button press immediately at the completion of the video within a two-second window. The key mapping for the two keys that coded for violent and nonviolent outcomes was counterbalanced (i.e. if one operator used the right button for the ‘violent’ response and the left button for the ‘non violent’ response another operator would use the left button for the ‘violent’ response and the right button for the ‘non violent’, the same applied to the novices). Participants pressed the response button with their dominant hand. After the completion of the video there was 2 seconds before the next trial began. Each run ended with 10 seconds of fixation and 16 seconds of blank screen. The duration of each run was 3 minutes and 36 seconds.

Functional images

Scanning was performed on a Siemens 3T Tim Trio MRI scanner. The functional scan consisted of two runs for each experiment (TR = 2000 ms; TE = 30 ms; 32 Slices; 3 mm³ isovoxel; 70X70 image resolution; 138 Volumes).

Anatomical images

For each participant we acquired a high-resolution anatomical scan. The anatomical scans were T1 weighted MPRAGE sequences, and had a TR of 1900 ms and a TE of 2.52 ms. We collected 192 slices at a voxel resolution of 1 mm³, and image resolution of 256 X 256.

fMRI data analysis

The functional and anatomical images were analyzed using Brain Voyager QX 2.4.2.2070 (Brain Innovation, Maastricht, The Netherlands).

Pre-processing of functional data

Functional imaging data (DICOM format) were pre-processed by performing a slice scan time correction. This was performed using sinc interpolation based on information about the TR (2000 ms) and the order of slice scanning. 3D motion correction (6 df) was performed to detect and correct for small head movements, by spatial alignment of all volumes of a subject to the first volume by rigid body transformations. Estimated translation and rotation parameters were inspected and never exceeded 3 mm or 2°. Linear trend removal and

temporal high-pass filtering were then applied to remove low-frequency nonlinear drifts of three or fewer cycles (0.011 Hz) per time course. The functional MR images were smoothed using a Gaussian filter with fullwidth at half-maximum (FWHM) equal to 8 mm.

Pre-processing of anatomical data

The anatomical data (DICOM format) of each subject were loaded and converted into BrainVoyager's internal 'VMR' data format. The data were then resampled to 1 mm resolution and transformed into anterior commissure–posterior commissure (AC–PC) and Talairach standard space. The three spatial transformations were combined and applied backwards in one step to avoid quality loss due to successive data sampling. The two affine transformations, iso-voxel scaling and AC–PC transformation, were concatenated to form a single 4×4 transformation matrix. For each voxel coordinate in the target (Talairach) space a piecewise affine 'Un-Talairach' step was performed, followed by application of the inverted spatial transformation of the aforementioned matrix. The computed coordinates were used to sample the data points in the original 3D space using sinc interpolation.

Normalization of functional data

To transform the functional data into Talairach space, the functional time series data of each subject was first coregistered with the subject's 3D anatomical dataset, followed by the application of the same transformation steps as performed for the 3D anatomical dataset (see above). This step resulted in normalized 4D volume time course ('VTC') data. In order to avoid quality loss due to successive data sampling, normalization was performed in a single step combining a functional–anatomical affine transformation matrix, a rigid-body AC–PC transformation matrix, and a piecewise affine Talairach grid scaling step. As described for the anatomical normalization procedure, these steps were performed backwards, starting with a voxel in Talairach space and sampling the corresponding data in the original functional space. The functional slices were coregistered to the anatomical volume using manual alignment to obtain optimal fit and reduce as much as possible the geometrical distortions of the echo-planar images. The necessary scaling adjustment was done interactively, using the appropriate transformation and visualization tools of BrainVoyager QX.

Analysis

First level analysis

Analyses were performed on the data of individual participants using multiple linear regression of the BOLD-response time course in each voxel, using four predictors (Display Type: Fight, Confront, Play and Nothing) when performing the first whole brain analysis, two predictors (Participants' responses: Fight, No fight) when performing the second whole brain analysis, and one predictor (saliency) defined by one TR when performing the third whole brain analysis. For each run of each participant's event-related data, a BrainVoyager protocol file (PRT) was derived, representing the onset and duration of the events for the different conditions. Hence, PRT files were created separately for each of the three analyses and for each participant. Predictors' time courses were adjusted for the hemodynamic response delay by convolution with a hemodynamic response function.

Second level analysis

For the analysis with the predetermined four display categories, we carried out a 2 (experience: novices and operators) X 4 (displays: fight, confront, play, and nothing) analysis of variance with Participant Experience as a between-participants factor and Display as a within-participants factor. For the analysis with the participants' judgements we carried out a 2 (experience: novices and operators) X 2 (judgements: fight, no fight) analysis of variance with Participant Experience as a between-participants factor and Judgements as a within-participants factor. For the analysis with the most salient event in each display we carried out a 2 (experience: novices and operators) X 2 (saliency: salient event, baseline) analysis of variance with Participant Experience as a between-participants factor and Saliency as a within-participants factor. The activations are reported for all three analyses at a threshold of $P < 0.001$ (uncorrected), and were corrected for multiple comparisons using the cluster-size threshold of $P < 0.05$ (for details see Goebel et al., 2006), based on a 3D extension of the randomization procedure described in Forman et al. (1995). In this method, for each statistical map the uncorrected voxel-level threshold was set at $P < 0.001$, and then was submitted to a whole-brain correction criterion based on the estimate of the map's spatial smoothness and on an iterative procedure (Monte Carlo simulation) for estimating cluster-level false-positive rates (i.e. the theoretical number of “false” positive voxels that are activated in each random map). After 1000 iterations the minimum cluster-size that yielded a cluster-level false- positive rate of 5% was used to threshold the statistical map. The minimum cluster-size for $P < 0.05$ is reported in voxels and in mm. The minimum cluster threshold that yielded a cluster-level false-positive rate of 5% was $k=3$, 81 mm³ for all the assessed maps. Finally, we examined the relationship between the found brain activation (i.e.,

averaged betas weights) derived from each cluster for each participant and the participants' rating of empathic concern and perspective taking, by carrying out a series of linear regression analyses. Additionally, we carried out linear regression analyses to examine the relationship between the brain activation and the years of CCTV experience.

Intersubject correlation analysis

We used intersubject correlation (Hasson et al., 2004) to examine how the brain activity of the two different groups synchronised when watching the 16 second movies. To achieve this we correlated the timecourse of each voxel in the brain for each CCTV operator with each other operator and for each novice with each other novice while were watching the same clips. This enabled us to exclude the subject's own data in the obtained average (Cantlon and Li, 2013). After obtaining correlation maps for each participant we then calculated a mean image across the paired maps for each operator and each novice. Each operator and each novice thus had one map of neural correlation representing their mean similarity to their own group. We then performed a whole brain analysis comparing operator-to-operator intersubject correlation with novice-to-novice intersubject correlation. This gave us a measure of where in the brain the neural responses among operators were more similar (more synchronised as consequence of CCTV experience) than those among novices (for a similar analysis see Cantlon and Li, 2013).

Behavioural Results

Eye-tracking results before scanning

Thirty-six clips were used in the behavioral phase that preceded the fMRI study. Of these 36 clips, 9 were labelled as 'Fight' because they showed a sequence of behaviors leading up to a violent incident (although the violent incident itself was not shown and occurred after the clip ended); 9 were labelled as 'Confront' because the clips showed a sequence of behaviors similar to each fight clip, although no violent/harmful incident occurred after the clip ended; 9 were labelled as 'Play' because the clips showed people interacting in a playful manner; and finally 9 were labelled as 'Nothing' because the clips showed a variety of scenes where no violent/harmful behavior occurred, and they were taken from similar locations and with similar camera views to the other clip types. Eye movements were tracked for 21 participants (11 CCTV operators and 10 novices) to examine which visual cues were most informative for accurately predicting violence, and whether there were differences in gazing time or fixation duration for the operators compared with the novices. For technical reasons, the data of one

of the operators had to be excluded from the analysis, leaving a total of 20 participants (10 CCTV operators and 10 novices).

Gazing time was defined as the percentage of time participants spent fixating, in relation to the overall clip length. The mean gazing time for each participant in each Clip Category was calculated and was submitted to a two-factor ANOVA (experience x clip); this revealed no main effect of Participant Experience ($F(1, 18) = 1.897, p = 0.185$), a main effect of Clip Category ($F(3, 54) = 35.349, p < 0.001$), and a significant interaction between these two factors ($F(3, 54) = 2.787, p = 0.049$). The fact that no main effect of Participant Experience was found indicated that the CCTV operators spent a similar percentage of time making fixations ($M = 76.16, S.D. = 4.19$) as did novice participants ($M = 78.50, S.D. = 4.80$). The main effect of Clip Category indicates that there were significant differences in participants' gazing times when viewing different types of clips. This main effect was further investigated with paired samples t-tests, after applying a Bonferroni correction for multiple comparisons. Participants exhibited significantly longer gazing times for clips in the matched confront clip category ($M = 80.08, S.D. = 3.66$), than for fight clips ($t(19) = -3.588, p = 0.008; M = 78.31, S.D. = 3.99$), to play clips ($t(19) = 8.409, p < 0.001; M = 74.54, S.D. = 4.78$) or nothing clips ($t(19) = 6.521, p < 0.001; M = 76.37, S.D. = 4.34$). Similarly, significantly longer gazing times were found for fight clips than play clips ($t(19) = 5.465, p < 0.001$), or nothing clips ($t(19) = -3.818, p < 0.001$). Finally, play clips elicited longer gazing times than nothing clips ($t(19) = -2.905, p = 0.036$). The interaction was due to the shorter amount of time CCTV operators spent gazing at fight clips (operators: $M = 76.58, S.D. = 3.18$, novices: $M = 80.05, S.D. = 4.10$), confront clips (operators: $M = 78.51, S.D. = 3.19$, novices: $M = 81.64, S.D. = 3.55$), and nothing clips (operators: $M = 75.23, S.D. = 4.95$, novices: $M = 77.53, S.D. = 3.52$) than novices, while no difference between the two groups was evident when gazing at the play clips (operators: $M = 74.31, S.D. = 4.51$, novices: $M = 74.78, S.D. = 5.28$).

The mean fixation duration data were analyzed using a two-factor (Experience and Clip Category) ANOVA that revealed once again no main effect of Participant Experience ($F(1, 18) = 1.517, p = 0.234$), a main effect of Clip Category ($F(3, 54) = 8.025, p < 0.001$), and no interaction between these two factors ($F(3, 54) = 1.484, p = 0.229$). Hence, CCTV operators and novices had similar mean fixation durations (operators: $M = 0.34, S.D. = 0.02$; novices: $M = 0.36, S.D. = 0.04$). The main effect of Clip Category was further investigated with paired samples t-tests, after applying a Bonferroni correction for multiple comparisons. Participants exhibited significantly longer fixation times for clips in the matched confront clip category ($M = 0.36, S.D. = 0.03$), than play clips ($t(19) = 3.672, p = 0.008; M = 0.34, S.D. = 0.02$), or

nothing clips ($t(19) = 3.362, p = 0.012; M = 0.34, S.D. = 0.2$), while no difference was found between confront and fight clips ($t(19) = -2.178, p = 0.168; M = 0.35, S.D. = 0.03$). No other significant differences were found when comparing the mean fixation times between the remaining clip categories. In summary, the fixation analyses failed to show any significant differences between CCTV operators and novices.

Selection of salient events in the clips

For the selected clips another series of analyses was carried out on the eye-tracking data to examine whether there was any difference between the events that CCTV operators and novices found most salient in the clips. To this end we examined the most salient event in terms of both the variable *time* (i.e. the point in time when participants focused on the spatial position) and *density* (i.e. reflecting the level of agreement in space). We performed a 2 (expertise: CCTV operators, novices) x 4 (clip types: fight, confront, play, and nothing) ANOVA with expertise as a between-subjects factor and clip types as a within-subjects factor. The results for *time* showed no main effect for the clip types ($F(3,48) = 1.792, p = 0.195$) and no significant interaction between clip types and experience ($F(3,48) = 1.214, p = 0.341$). Also there was no significant effect for CCTV experience, i.e. no significant difference between the operators and controls ($F(1,48) = 0.194, p = 0.665$). We found similar results for *density*, with no main effect for the clip types ($F(3,48) = 0.545, p = 0.659$) and no significant interaction between clip types and experience ($F(3,48) = 0.735, p = 0.548$). Similarly to before, there was no significant effect for CCTV experience, i.e. no significant difference between the operators and controls ($F(1,48) = 2.017, p = 0.112$). Therefore, for both operators and novices, the event perceived as being the most salient occurred at the same time into the 16-second display and at a similar spatial position. Due to the lack of difference between CCTV operators and novices we considered the salient event in the clip to be the one at which we found the highest level of agreement for all participants (i.e. the highest density in x-y gaze coordinates and time). These salient events from the behavioural data were used subsequently in the fMRI experiment to define the time at which a salient event occurred during each of the 16-second displays. The practice of using group data obtained outside the scanner has been advocated previously (Zacks, et al., 2006) for defining times when similar data from the scanning session is less reliable, as is our case due to the limited eye movement data obtained during scanning.

Behavioral results before scanning

After each clip was shown during the eye-tracking experiment, the same 21 participants were asked to rate the likelihood that a violent incident occurred after the clip ended, using a six-point scale that ranged from 1 (extremely unlikely) to 6 (extremely likely). A 2 x 4 factor repeated measures analysis of variance (ANOVA) showed a main effect of Participant Experience ($F(1, 19) = 7.111, p = .015$) and of Clip Category ($F(3, 57) = 136.954, p < .001$), with no interaction between these factors ($F(3, 57) = 1.023, p = 0.389$). The CCTV operators rated the clips as being likely to end in violence more often than the novices did. Regarding the display categories, pairwise comparisons (conducted after applying a Bonferroni correction at $p < 0.05$) indicated that whereas fight and confront clips received similar average ratings of 'fight' responses (i.e. 3.96 and 4.16 out of 6 respectively), fight clips received significantly higher ratings of 'fight' responses than play clips (i.e. 2.37), or nothing clips (i.e. 1.79). Similarly, the confront clips received significantly higher ratings of 'fight' responses than play or nothing clips. Finally, play clips received significantly higher ratings of 'fight' responses than nothing clips.

A signal detection analysis (Green and Swets, 1966) was applied to the likelihood ratings; this involved recoding rating responses of 4 and greater as predicting a violent outcome, and rating responses of 3 and less as not predicting a violent outcome. Judging a fight display to have a violent outcome was scored as a 'hit', and judging a confront, play or nothing clip as not having a violent outcome was scored as a 'correct rejection'. A calculation of sensitivity index, d' , and bias, C , was performed for each participant and produced an average d' of 1.067 and bias C of -0.035 for operators and an average d' of 0.782 and bias C of 0.257 for novices. A comparison of means using Mann-Whitney tests showed a marginally significant difference in d' between the two groups ($Z = -1.480, p = 0.051$, one-tailed). Similar analyses carried out on the criterion data showed a significant difference in C between the two groups ($Z = -1.799, p = 0.034$, one-tailed). This meant that operators were more sensitive to situations that would end in violence, but were also more biased than novices in their predictions of violent outcome when viewing a clip.

Additionally, when the eye-tracking part of the experiment was complete, participants viewed all the clips a second time, and immediately after each clip they were asked to rate the perceived level of hostility using a ten-point rating scale. This scale ranged from 1 (very jovial, no hostility), through 5/6 (neutral, benign, no emotions present), to 10 (very hostile, aggressive). Participant ratings of hostility were analyzed using a 2 x 4 factor repeated measures ANOVA. The data were analyzed along two different factors, Participant Experience (CCTV operator or Novice participant) and Clip Type (Fight, Confront, Play and

Nothing). This ANOVA revealed no main effect of Participant Experience ($F(1, 19) = 1.159$, $p = 0.295$), a main effect of Clip Category ($F(3, 57) = 218.831$, $p < 0.001$), and no interaction between these two factors ($F(3, 57) = 0.471$, $p = 0.704$). The main effect of Clip Category indicates that there were significant differences in participants' hostility ratings for different types of clips. This main effect of Clip Category was further investigated, again via pairwise comparisons using Bonferroni correction. Participants gave significantly higher hostility ratings for clips in the confront clip category (i.e. 7.88 out of 10) than fight clips (i.e. 7.09 out of 10), play clips (i.e. 3.65), or nothing clips (i.e. 4.02). Similarly, a significant difference was found in the hostility ratings between fight clips and play clips, and between fight clips and nothing clips. No significant difference in ratings of hostility between play clips and neutral clips was observed. The result that levels of hostility were greater for the confront than the fight displays is useful to rule out that participants could use a simple heuristic based on how much aggression was displayed to make their judgement of whether violence would or would not occur.

Behavioral results during scanning

Using the results of the behavioral and eye tracking experiments (see above and Methods section), 24 CCTV clips of 16 seconds were selected to be used during the fMRI scan. Of these 24 clips, 6 were a subset of the 'fight' clips, 6 of the 'confront', 6 of the 'play', and finally 6 of the 'nothing' clips. Thirty individuals participated in the fMRI experiment: 15 CCTV operators and 15 individuals with no experience of using CCTV surveillance to monitor human behavior in their workplace. None of these participants had seen the clips before.

While lying in the scanner, participants were asked to report the likelihood of a violent event occurring after the end of each clip, by pressing two buttons (one for 'fight will occur' and one for 'no fight will occur'). The numbers of correct responses for 15 operators and 14 novices (the behavioral data for one of the novices was lost due to a technical problem) were then analyzed by carrying out an ANOVA using the factors of Participant Experience (novice, operator) x Clip Category ('fight', 'confront', 'play', 'nothing'). The results of this analysis showed no main effect of Participant Experience ($F(1, 27) = 0.03$, $p = 0.864$), an effect of Clip Category ($F(3, 81) = 20.307$, $p < 0.001$), and a significant interaction ($F(3, 81) = 3.019$, $p = 0.035$). Repeated simple contrast measures analysis indicated that the main effect of Clip Category was driven by the difference in correct responses received by fight and confront displays ($F(1, 27) = 30.131$, $p < 0.001$). No differences between the number of correct

responses for fight and play displays ($F(1, 27) = 0.103, p = 0.750$), or fight and nothing displays ($F(1, 27) = 1.07, p = 0.310$) were found. The interaction was also driven by a difference in number of correct responses given by operators and novices to fight and confront displays ($F(1, 27) = 6.565, p = 0.016$). Whereas operators gave a higher number of correct responses to the clips effectively ending in a fight than novices, novices gave a higher number of correct responses to confront clips than operators. In other words, operators predicted more violent outcomes than novices when the clips actually ended in violence, but at the same time they also predicted more violent outcomes than novices when the clips containing confrontational behavior did not end in violence. This finding is supported by the signal detection analysis (Green and Swets, 1966) applied to participants' responses. Judging a fight display to have a violent outcome was scored as a 'hit' and judging a confront, play or nothing display as not having a violent outcome was scored as a 'correct rejection'.

Calculation of sensitivity index, d' , and bias, C , was performed for each participant and produced an average d' of 1.671 and bias C of -0.667 for operators and an average d' of 0.977 and bias C of -0.098 for novices (Figure 1). A comparison of means using Mann-Whitney tests showed a significant difference in d' between the two groups ($Z = -2.117, p = 0.034$, two-tailed). Similar analyses carried out on the criterion data showed a significant difference in C between the two groups ($Z = -2.008, p = 0.045$, two-tailed). We report the results only for the non-parametric analysis because the d' values for the novices group were not normally distributed, as shown by the Kolmogorov–Smirnov test of normality ($p = .003$). This confirmed that CCTV operators were more sensitive to situations with a real violent outcome, but also more biased towards predicting a violent outcome when no real violent outcome was present. These findings replicate those obtained with the different group of CCTV operators and novices involved in the behavioral phase before scanning.

----- Fig. 1 -----

fMRI Results

fMRI results for clip types

Here we explore the effect of the different types of clips on novice and operator brain responses, irrespective of whether participants perceived the clips as ending in violence or not. A 2 way ANOVA with Participant Experience as a between-subjects factor and Clip Category as a within-subjects factor identified two regions - the right parahippocampal gyrus and the right uncus (a region located in the anterior part of the parahippocampal gyrus) -

which exhibited significant differences for level of experience, and one region - the left superior frontal gyrus (SFG) - which exhibited an interaction between Participant Experience and Clip Category. Examination of the contrast parameters changes (Fig. 2, left panel) in the parahippocampal gyrus and in the uncus showed that the effect of experience was driven by a consistent reduction of activity in these areas for the operators when compared to the novice group. In the right uncus, a linear regression analysis identified a significant correlation between neural activity when viewing the confront clips and operators' years of experience ($r = 0.525$, $p = 0.044$). Deviation tests of within-subjects contrast (with fight as reference category) indicated that the significant interaction in SFG was driven by higher activity in this area for the operators than the novices for the play clips ($F(1, 28) = 23.783$, $p < 0.001$). This result was further supported by independent samples test analyses (fight: $t(28) = 1.292$, $p = 0.207$; confront: $t(28) = 1.004$, $p = 0.324$; play: $t(28) = 2.625$, $p = 0.014$; nothing: $t(28) = 0.742$, $p = 0.464$). In this same area, a linear regression analysis identified a significant correlation ($r = 0.545$, $p = 0.036$) between neural activity for the play clips and operators' indices of perspective taking (see Fig. 2). Therefore, these findings indicate that CCTV experience results in a reduction of activity in parahippocampal areas regardless of clip type, and in an increase of activity in SFG specific to play clips. However, using pre-defined experimental labels does not reflect participants' perceptions of intent in the viewed CCTV clips. Indeed, not all of the fight clips were judged as ending in violence, while many of the confront clips and some of the play and nothing clips were judged as ending in violence. For this reason, in the next section we describe the fMRI results based on participants' predictions of whether a situation would or would not result in violence.

----- Fig. 2 -----

fMRI results for judged intent

Here we used the responses given by the participants during the scan to explore the differences between the novices and operators when either they did or did not predict violence (Intent judgements), no matter whether the clip ended up with a fight or not. A two-way ANOVA (Participant Experience X Intent Judgements) revealed a main effect of experience in the right uncus, similar to the first analysis. However, no effect of higher-order interaction was found. The anatomical location and details of the uncus activation are described in Fig. 3. Changes in contrast estimates indicate that the main effect of experience was attributable to a lower level of activation in CCTV operators relative to novices in the uncus when viewing the clips. A positive correlation ($r = 0.529$, $p = 0.043$) was found

between operators' contrast estimates in the uncus when viewing the CCTV clips they judged to end in fights and their years of experience.

----- **Fig. 3** -----

fMRI results for salient events

The preceding results indicate that CCTV experience selectively reduces neural activity in areas of the parahippocampal gyrus. In order to understand whether this functional difference between the two groups in the parahippocampal gyrus was linked to what participants considered as being relevant information in the clips, we carried out a further analysis. In this third whole-brain analysis, we examined the brain activity of our current set of novice and operator observers at the time of a particularly salient event in the clip. This was carried out during the eye-tracking phase preceding the fMRI (see above). We measured the variability of the eye movements to find the time and spatial location when the observers' eye gazes were most in agreement with each other. This was termed the 'salience measure', since we assumed that when observers were all looking at the same visual point at the same time, this would correspond to a salient event in the clip. A 2 way ANOVA, with Participant Experience as a between-subjects factor and Saliency as a within-subjects factor, identified five regions that exhibited significant differences for level of experience: the bilateral parahippocampal gyrus, the right inferior temporal gyrus (ITG), the right superior occipital gyrus (SOG), and the left inferior occipital gyrus (IOG). Examination of the contrast parameter changes (Fig. 4) showed that the effect of experience was driven by a consistent reduction in activity for the operator group when compared to the novice group in all these areas. The presence, once again, of the parahippocampal areas among the regions identified in this third analysis confirms that these areas were strongly affected by CCTV experience, and by the way the two groups used salient information in the CCTV footage.

----- **Fig. 4** -----

Intersubject correlation analysis

We performed a whole brain analysis comparing operator-to-operator intersubject correlation with novice-to-novice intersubject correlation by performing an independent samples t-test over the mean correlation maps of individual subject (for a similar analysis see Cantlon and Li, 2013). Operators showed significantly higher intersubject correlation than novices in five brain regions (Figure 5): bilateral anterior superior temporal gyrus (BA 38), ventral striatum,

left middle temporal gyrus (BA 39), and left inferior parietal lobule (BA 39). On the contrary, novices showed higher intersubject correlation than operators in widespread regions of temporal, parietal and occipital cortex (Figure 5). Hence, in distinction to the many regions found in novices, the similarity between operators' neural responses is focused in a few specific regions.

----- Fig. 5 -----

Discussion

In the present study we used an fMRI paradigm to investigate how experience with CCTV footage alters behavioral and neural correlates when predicting violent episodes. Our findings demonstrate that exposure to CCTV footage leads to a consistent reduction in brain activity in regions of the parahippocampal gyrus. This effect was not specific to clips where some sort of confrontation between people was displayed, but was generalized to all CCTV clips. Nevertheless, the years of CCTV experience and perspective taking modulated the reduction of activity in different brain regions and for specific CCTV clips, further reinforcing the idea that the observed changes in level of brain functions are a consequence of changes in learning and/or habituation. The greater ability of operators to rightly predict a violent outcome based on CCTV footage indicates that experience and training makes these individuals better at judging intent to harm based solely on visual experience and observation of people's behaviour. Our results support recent findings from studies comparing expert kickers' and goalkeepers' judgements of deceptive actions, which showed that a better ability to detect deceptive actions came from visual rather than motor expertise (Tomeo et al., 2012). Nevertheless, CCTV operators were also more biased than novices in predicting a violent outcome when people acted in a confrontational manner, which could also be a consequence of the found reduction in brain activity. Notably, the magnitude of learning-induced deactivation in the uncus for the operator group decreased with years of CCTV experience, indicating that the greatest reduction in brain activity occurs at the beginning of training and practice. This excludes the explanation that this lower level of activity in the operator group was caused simply by the degree of overexposure to CCTV footage. In other words, if the found reduction in activity was caused only by the degree of overexposure, then we would have expected further brain activity reduction during the following years of CCTV experience. Furthermore, the overall tendency of CCTV operators to give lower ratings of perspective taking and empathic concern than novices, and to have reduced blood oxygenation level dependent (BOLD) signals when fixating salient events in the clips,

suggests that this may be a consequence of a detached perspective learnt by the CCTV operators throughout their training and practice. The evident reduction in synchronisation of neural responses across the brain for operators compared to novices further support the idea that CCTV experience results in more specific and efficient brain processes when predicting violent behaviour from naturalistic visual scenes.

Model-based analysis

Superior frontal gyrus

When we explored the effect of the different types of clips on novice and operator brain responses, we found, irrespective of participants' responses, an effect of experience on the right parahippocampal gyrus, the right uncus and the left superior frontal gyrus (SFG). This latter region was the only one showing greater activation for operators than for novices, and this greater activity was confined to the play clips. Additionally, the contrast estimates obtained for the SFG from individual CCTV operators viewing the play clips correlated with their scores of perspective taking, whilst this was not the case for the novice group. This difference between CCTV operators and novices could be explained by the operators' tendency to look closely at the play clips, as some of the operators reported that playful interactions can unexpectedly result in fights later on. In other words, making the right prediction when watching play clips is more challenging for operators than novices (who mostly perceive the play clips as harmless). SFG is known to be important for positive and negative emotion regulation (Goldin et al., 2008; Mak et al., 2009; e.g. Ochsner et al., 2004; Urry et al., 2006), in that, for example, it shows greater activation when participants consciously increase or decrease negative emotions (e.g. Mak et al., 2009; Ochsner et al., 2004). Emotional self-regulation is particularly important for CCTV operators when they are making an objective decision on whether a situation will result in violence and thus require intervention. Clips containing either clear negative confrontational behaviour (i.e. fight and confront clips) or no confrontational behaviour at all (i.e. nothing clips) are less ambivalent for CCTV operators in their emotional meaning, and hence do not require as much emotional self-regulation and perspective-taking as the play clips. As we did not use any screening method to assess the level of emotional self-regulation in novices and operators we cannot make any strong conclusion on whether the found group difference in SFG truly reflects a difference in emotional self-regulation processes. Hence, further studies could employ methods to characterize emotional self-regulation together with neuroimaging techniques to clarify the involvement of SFG in the prediction of others' harmful behaviour.

Parahippocampal gyrus

Differences in activity in the parahippocampal areas as well as superior frontal regions have been previously reported between expert physicians and non-experts (e.g. Cheng et al., 2007), while they watched videos showing acupuncture being used on the mouths, hands, and feet of patients. In a more recent paper Petrini et al. (2011) reported a reduction in parahippocampal activity in a group of expert drummers when they watched synchronised audio-visual music displays, but this effect was not found in a group of novices. Thus, long-term experience and familiarity with the portrayed actions and situations seems to be at the root of reported deactivation in this area. Strong support for this idea comes from an elegant study by Schiffer et al. (2013) in which the authors show how the association between events (actions) within a clip modulate the activity of the parahippocampal gyrus. Schiffer et al. (2013) based their study on the idea that the more frequently one particular action follows another action, the more likely they will be associated in a strong internal model. They showed that when observers had a strong internal model there was greater activity in the right parahippocampal gyrus when participants were exposed to action sequences that began identically to the original sequence but contained actions that diverged from the model's expectations. The consistent reduction in parahippocampal gyrus activity that we found in CCTV operators may thus reflect the higher ability of this group to predict how the CCTV clip would end based on their greater exposure to similar associated actions. That the parahippocampal gyrus might show greater involvement when an action is recognised based on the portrayed visual information is supported by previous studies showing greater activity in this area for action recognition than action imitation (e.g. Decety et al., 1997). The results of our fMRI analysis for the salient events in the clips support this idea by showing a clear effect of experience on the parahippocampal gyrus when comparing the brain activity of CCTV operators to that of novices for the most visually salient events in the clips. It is important to remember that no differences between the two groups were found when selecting the salient events. This excludes the explanation that any found effect of experience is the result of differences in what the two groups of participants found salient in the clips when predicting the occurrence of a violent event. Effects of CCTV experience were detected in regions of the bilateral parahippocampal gyrus, and in a group of temporo-occipital regions, but not in the uncus.

The posterior parahippocampal regions in rhesus monkeys receive projection from visual areas BA 19 and 18 (Van Hoesen, 1982), and similar connectivity in humans may explain why we detected these visual areas when analysing the fMRI data for visual saliency.

CCTV operators do not need to activate as much as novices when the relevant visual information is gathered by the temporo-occipital regions and then sent to the parahippocampal gyrus. Lesion studies in monkeys have shown that lesions of the perirhinal and parahippocampal cortex are at the root of memory impairment (e.g. Zola-Morgan et al., 1989), and lesion studies in humans with pharmacoresistant medial temporal lobe epilepsy have confirmed the unique role of the parahippocampal gyrus for associative learning and recall of objects and faces (e.g. Weniger et al., 2004). Because the people and scenes portrayed in the CCTV footage change continually, the reduction of activity found in CCTV operators cannot reflect acquired familiarity with specific visual information, but rather must reflect familiarization with portrayed actions and scenes (Nakamura et al., 2000). The idea that the parahippocampal gyrus is implicated in action observation and recognition is suggested by previous studies showing greater activity in this area for action recognition than action imitation, and for meaningful than meaningless actions (Decety et al., 1997). In a study where single-neuron responses were recorded in patients during the execution and observation of actions, Mukamel et al. (2010) showed that a subset of neurons in the parahippocampal gyrus responded specifically to action observation, another subset specifically to action execution, and a smaller subset to both observation and execution. Further to this, Mukamel et al. (2010) showed that some of the neurons responding to both observation and execution of the same action decreased their firing rate during both conditions, while some of the other neurons either increased their firing rate during execution and decreased during observation or the other way around. Hence the parahippocampal gyrus appears to contain a high number of neurons that process different aspects of action recognition and execution, and they are probably called into play when familiar memories of the observed and/or performed action are accessed. Our results show that having greater visual experience than motor experience/familiarity with others' actions can significantly decrease activity bilaterally in the parahippocampal gyrus, thus supporting the suggestion that this area may be part of an extended "mirror neuron" mechanism or an "action recognition" network (Mukamel et al., 2010).

Uncus

Seeger and Cincotta (2006) found a reduction in activity during the rule-learning task in a hippocampal region very similar to our uncus. In line with the results of Seeger and Cincotta (2006), we found an opposite trend in activity in the frontal and hippocampal areas: while activity in the frontal areas increased for the operator group when compared to the novices,

the hippocampal activity declined. The activity we found in the right uncus when analyzing the effect of experience for different display types was still present when analyzing the brain activity based on participants' responses. Exactly as in the first analysis, the level of activity in this area was found to increase with years of CCTV experience for the displays judged as leading to a violent act, but no such correlation was found for displays judged as non-violent. This finding extends and clarifies the results of the previous analysis, in which years of CCTV experience were found to positively correlate with brain activity elicited by confront displays, for which operators, more than novices, were found to be biased towards predicting a violent outcome. That is, greater activity in the uncus reflects greater CCTV experience in predicting a violent outcome. The significant inverse correlation we found between years of CCTV experience and operators' rating of empathic concern suggests that initially CCTV operators may strongly suppress their emotional involvement, with a consequent reduction in the uncus activity. After being repeatedly exposed to violent events, CCTV operators become emotionally detached from (less empathic towards) the portrayed situations and thus do not need to consciously suppress their emotions any more, which results in a gradual reactivation of the uncus. The uncus anatomical location, together with its described function, supports this possibility. Indeed, the uncus is a transitional area located in the anterior part of the parahippocampal gyrus (Shah et al., 2012), overlooking the amygdala. Activity in this area has often been associated with conscious subjective experiences of fear (Bancaud, 1987; Vuilleumier et al., 2001), with automatic and unconscious recognition of fearful faces (de Gelder et al., 2005) and with activity evoked by unmasked threat words (Naccache et al., 2005). Thus reduced activity in this area for the CCTV operators compared to the novices may reflect desensitization to violent/fearful situations in response to the initial training and practice, and may stem from the same emotional regulation mechanism causing activity in SFG to increase for operators but not for novices.

Intersubject correlation analysis

A whole brain analysis comparing operator-to-operator intersubject correlation with novice-to-novice intersubject correlation revealed significantly higher operators' intersubject correlation than novices in five brain regions: bilateral anterior superior temporal gyrus (BA 38), ventral striatum, left middle temporal gyrus (BA 39), and left inferior parietal lobule (BA 39). The superior temporal gyrus is known to activate when people observe others' actions (Decety and Grezes, 1999). Activation in the anterior part of the superior temporal gyrus (BA 38) has been often found in response to social features (e.g. Lahnakoski et al., 2012) and

others' angry and threatening emotional gestures (Grezes et al., 2007; Grosbras and Paus, 2006; Pichon et al., 2009). For example, Pichon et al. (2009) showed greater activation in both the anterior part of the superior temporal gyrus and the temporal pole when contrasting brain responses to actions expressing anger with those expressing fear. Similarly, the ventral striatum has been associated to emotional and social judgements (Blood and Zatorre, 2001; Dapretto et al., 2006; Kampe et al., 2001), and its activity has been found to be greater for observed martial actions than dance actions (Porges and Decety, 2013). A peak of activation with very similar Talairach coordinates to our middle temporal gyrus (-42, -60, 30) has also been found by Grezes et al. (2004) when subjects judged a displayed person to have deceptive intentions. Finally, neurons in the inferior parietal lobule have been shown to be able to code observed actions as well as to represent the intention of such actions (Buccino et al., 2001; Fogassi et al., 2005; Grezes and Decety, 2001). The inferior parietal lobule is involved in the sense of agency and has a function in perspective taking (Shane et al., 2009), and its left part activates more strongly than the right when an observer simulates and reproduces the action of another (Chaminade and Decety, 2002; Decety et al., 2002; Jackson et al., 2006; Ruby and Decety, 2001). Hence, long-term experience with judging intent to harm in others' increases the synchronisation in neural responses in regions known to be important for recognising actions, the intention of actions and their emotional significance.

Conclusion

In conclusion, our findings suggest that experience with CCTV footage modifies neural response in several regions of the brain. CCTV experience appears to reduce brain activity in action recognition areas (parahippocampal gyrus) and either to reduce or increase activity in areas involved in emotion regulation (e.g. uncus and SFG). Although the parahippocampal gyrus consistently showed reduced activity in CCTV operators compared to novices irrespective of the analysis carried out, the reduction in activity in its anterior (the uncus) and posterior part seemed to have different causes. Whereas the direction of effect of experience on the uncus changed with years of experience, this was not the case for the posterior part of the parahippocampal gyrus. CCTV experience also increases the synchronisation of neural responses in fewer regions of the brain (i.e. bilateral anterior superior temporal gyrus, left middle temporal gyrus, left ventral striatum and left inferior parietal lobule) than for novices. Moreover, these particular areas are known to be important for recognising actions, the intention of actions and their emotional significance.

Altogether these findings indicate that CCTV experience results in more efficient neural processes developed to improve the prediction of harmful behaviour from complex naturalistic visual scenes. Our findings, besides demonstrating how visual experience with complex social scenes changes brain functions, make an important contribution to understanding how individuals involved in security and defense make judgements of harmful intention. This is an area where it is recognised that there is a shortage of empirical knowledge (Weinberger, 2010).

Acknowledgements

This work was funded by the Human Dimension and Medical Research Domain of the Dstl Programme Office for the UK Ministry of Defence. KP contribution was partially supported by a grant from the ESRC (RES-060-25-0010). We also would like to thank Andrea MacFadyen, Kirsty Wright, Christina Bannerman, Jesus Rodriguez-Perez and Katie Noble for their assistance in running the study. Finally we thank all the CCTV staff and other participants for their contribution to this research.

References

- Amodio, D.M. and Frith, C.D. Meeting of minds: the medial frontal cortex and social cognition. *Nature reviews. Neuroscience* **7** (4): 268-277, 2006.
- Baddeley, A., Emslie, H., and Nimmo-Smith, I. The Spot-the-Word test: a robust estimate of verbal intelligence based on lexical decision. *The British journal of clinical psychology / the British Psychological Society* **32** (Pt 1): 55-65, 1993.
- Bancaud, J. Clinical symptomatology of epileptic seizures of temporal origin. *Revue neurologique* **143**: 392-400, 1987.
- Berkowitz, A.L. and Ansari, D. Expertise-related deactivation of the right temporoparietal junction during musical improvisation. *NeuroImage* **49** (1): 712-719, 2010.
- Blood, A.J. and Zatorre, R.J. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences of the United States of America* **98** (20): 11818-11823, 2001.
- Brewer, J.A., Worhunsky, P.D., Gray, J.R., Tang, Y.Y., Weber, J., and Kober, H. Meditation experience is associated with differences in default mode network activity and connectivity. *Proceedings of the National Academy of Sciences of the United States of America* **108** (50): 20254-20259, 2011.
- Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., and Freund, H.J. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *The European journal of neuroscience* **13** (2): 400-404, 2001.
- Buss, A.H. and Perry, M. The aggression questionnaire. *Journal of personality and social psychology* **63** (3): 452-459, 1992.
- Calvo-Merino, B., Glaser, D.E., Grezes, J., Passingham, R.E., and Haggard, P. Action observation and acquired motor skills: an FMRI study with expert dancers. *Cereb Cortex* **15** (8): 1243-1249, 2005.
- Calvo-Merino, B., Grezes, J., Glaser, D.E., Passingham, R.E., and Haggard, P. Seeing or doing? Influence of visual and motor familiarity in action observation. *Current biology : CB* **16** (19): 1905-1910, 2006.
- Cantlon, J.F. and Li, R. Neural activity during natural viewing of Sesame Street statistically predicts test scores in early childhood. *PLoS biology* **11** (1): e1001462, 2013.
- Castelli, F., Frith, C., Happe, F., and Frith, U. Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain : a journal of neurology* **125** (Pt 8): 1839-1849, 2002.
- Chaminade, T. and Decety, J. Leader or follower? Involvement of the inferior parietal lobule in agency. *Neuroreport* **13** (15): 1975-1978, 2002.
- Cheng, Y., Lin, C.P., Liu, H.L., Hsu, Y.Y., Lim, K.E., Hung, D., and Decety, J. Expertise modulates the perception of pain in others. *Current biology : CB* **17** (19): 1708-1713, 2007.
- Cross, E.S., Hamilton, A.F., and Grafton, S.T. Building a motor simulation de novo: observation of dance by dancers. *NeuroImage* **31** (3): 1257-1267, 2006.
- Cross, E.S., Hamilton, A.F., Kraemer, D.J., Kelley, W.M., and Grafton, S.T. Dissociable substrates for body motion and physical experience in the human action observation network. *The European journal of neuroscience* **30** (7): 1383-1392, 2009.
- Dapretto, M., Davies, M.S., Pfeifer, J.H., Scott, A.A., Sigman, M., Bookheimer, S.Y., and Iacoboni, M. Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature neuroscience* **9** (1): 28-30, 2006.
- de Gelder, B., Morris, J.S., and Dolan, R.J. Unconscious fear influences emotional awareness of faces and voices. *Proceedings of the National Academy of Sciences of the United States of America* **102** (51): 18682-18687, 2005.
- Decety, J., Chaminade, T., Grezes, J., and Meltzoff, A.N. A PET exploration of the neural mechanisms involved in reciprocal imitation. *NeuroImage* **15** (1): 265-272, 2002.
- Decety, J. and Grezes, J. Neural mechanisms subserving the perception of human actions. *Trends in cognitive sciences* **3** (5): 172-178, 1999.

Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., Grassi, F., and Fazio, F. Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain : a journal of neurology* **120** (Pt 10): 1763-1777, 1997.

Diffley, C. and Wallace, E. CCTV: making it work: training practices for CCTV operators. *Home Office Policy Directorate Publication Number 9/98, Police Scientific Development Branch, Sandridge, 1998.*, 1998.

Epstein, R., Harris, A., Stanley, D., and Kanwisher, N. The parahippocampal place area: recognition, navigation, or encoding? *Neuron* **23** (1): 115-125, 1999.

Fadiga, L., Craighero, L., and Olivier, E. Human motor cortex excitability during the perception of others' action. *Current opinion in neurobiology* **15** (2): 213-218, 2005.

Fadiga, L., Fogassi, L., Pavesi, G., and Rizzolatti, G. Motor facilitation during action observation: a magnetic stimulation study. *Journal of neurophysiology* **73** (6): 2608-2611, 1995.

Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., and Rizzolatti, G. Parietal lobe: from action organization to intention understanding. *Science* **308** (5722): 662-667, 2005.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., and Noll, D.C. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med* **33** (5): 636-647, 1995.

Frith, C.D. and Frith, U. Interacting minds--a biological basis. *Science* **286** (5445): 1692-1695, 1999.

Frith, C.D. and Frith, U. The neural basis of mentalizing. *Neuron* **50** (4): 531-534, 2006.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. Action recognition in the premotor cortex. *Brain : a journal of neurology* **119** (Pt 2): 593-609, 1996.

Goebel, R., Esposito, F., and Formisano, E. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human brain mapping* **27** (5): 392-401, 2006.

Goldin, P.R., McRae, K., Ramel, W., and Gross, J.J. The neural bases of emotion regulation: reappraisal and suppression of negative emotion. *Biological psychiatry* **63** (6): 577-586, 2008.

Grafton, S.T. Embodied cognition and the simulation of action to understand others. *Annals of the New York Academy of Sciences* **1156**: 97-117, 2009.

Grafton, S.T., Arbib, M.A., Fadiga, L., and Rizzolatti, G. Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. *Experimental Brain Research* **112** (1): 103-111, 1996a.

Grafton, S.T., Fagg, A.H., Woods, R.P., and Arbib, M.A. Functional anatomy of pointing and grasping in humans. *Cereb Cortex* **6** (2): 226-237, 1996b.

Green, D.M. and Swets, J.A. *Signal detection theory and psychophysics*. Wiley: New York,, 1966.

Grezes, J. and Decety, J. Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Human brain mapping* **12** (1): 1-19, 2001.

Grezes, J., Frith, C., and Passingham, R.E. Brain mechanisms for inferring deceit in the actions of others. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **24** (24): 5500-5505, 2004.

Grezes, J., Pichon, S., and de Gelder, B. Perceiving fear in dynamic body expressions. *NeuroImage* **35** (2): 959-967, 2007.

Grosbras, M.H. and Paus, T. Brain networks involved in viewing angry hands or faces. *Cereb Cortex* **16** (8): 1087-1096, 2006.

Hasselmo, M.E. and Stern, C.E. Mechanisms underlying working memory for novel information. *Trends in cognitive sciences* **10** (11): 487-493, 2006.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. Intersubject synchronization of cortical activity during natural vision. *Science* **303** (5664): 1634-1640, 2004.

Herrington, J.D., Baron-Cohen, S., Wheelwright, S.J., Singh, K.D., Bullmore, E.T., Brammer, M., and Williams, S.C.R. The role of MT+/V5 during biological motion perception in Asperger Syndrome: An fMRI study. *Research in Autism Spectrum Disorders* **1**: 14-27, 2007.

Howard, C.J., Gilchrist, I.D., Troscianko, T., Behera, A., and Hogg, D.C. Task relevance predicts gaze in videos of real moving scenes. *Experimental Brain Research* **214** (1): 131-137, 2011a.

Howard, C.J., Pharaon, R.G., Korner, C., Smith, A.D., and Gilchrist, I.D. Visual search in the real world: evidence for the formation of distractor representations. *Perception* **40** (10): 1143-1153, 2011b.

Howard, C.J., Troscianko, T., Gilchrist, I.D., Behera, A., and Hogg, D.C. Searching for threat: factors determining performance during CCTV monitoring. In D. de Waard, J.G., F.L. Kooi, and K.A. Brookhuis (Eds.) (Ed.), *Human Factors, Security and Safety*: Maastricht, the Netherlands: Shaker Publishing, 2009.

Iacoboni, M., Koski, L.M., Brass, M., Bekkering, H., Woods, R.P., Dubeau, M.C., Mazziotta, J.C., and Rizzolatti, G. Reafferent copies of imitated actions in the right superior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America* **98** (24): 13995-13999, 2001.

Jackson, P.L., Meltzoff, A.N., and Decety, J. Neural circuits involved in imitation and perspective-taking. *NeuroImage* **31** (1): 429-439, 2006.

Kaiser, M.D., Hudac, C.M., Shultz, S., Lee, S.M., Cheung, C., Berken, A.M., Deen, B., Pitskel, N.B., Sugrue, D.R., Voos, A.C., Saulnier, C.A., Ventola, P., Wolf, J.M., Klin, A., Vander Wyk, B.C., and Pelphrey, K.A. Neural signatures of autism. *Proceedings of the National Academy of Sciences of the United States of America* **107** (49): 21223-21228, 2010.

Kampe, K.K., Frith, C.D., Dolan, R.J., and Frith, U. Reward value of attractiveness and gaze. *Nature* **413** (6856): 589, 2001.

Kelly, A.M. and Garavan, H. Human functional neuroimaging of brain changes associated with practice. *Cereb Cortex* **15** (8): 1089-1102, 2005.

Keysers, C. and Gazzola, V. Towards a unifying neural theory of social cognition. *Progress in brain research* **156**: 379-401, 2006.

Kirchhoff, B.A., Wagner, A.D., Maril, A., and Stern, C.E. Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **20** (16): 6173-6180, 2000.

Lahnakoski, J.M., Glerean, E., Salmi, J., Jaaskelainen, I.P., Sams, M., Hari, R., and Nummenmaa, L. Naturalistic fMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. *Frontiers in human neuroscience* **6**: 233, 2012.

Mak, A.K., Hu, Z.G., Zhang, J.X., Xiao, Z.W., and Lee, T.M. Neural correlates of regulation of positive and negative emotions: an fmri study. *Neuroscience letters* **457** (2): 101-106, 2009.

Mathiak, K. and Weber, R. Toward brain correlates of natural behavior: fMRI during violent video games. *Human brain mapping* **27** (12): 948-956, 2006.

Mitchell, J.P. Mentalizing and Marr: an information processing approach to the study of social cognition. *Brain research* **1079** (1): 66-75, 2006.

Mukamel, R., Ekstrom, A.D., Kaplan, J., Iacoboni, M., and Fried, I. Single-neuron responses in humans during execution and observation of actions. *Current biology : CB* **20** (8): 750-756, 2010.

Naccache, L., Gaillard, R., Adam, C., Hasboun, D., Clemenceau, S., Baulac, M., Dehaene, S., and Cohen, L. A direct intracranial record of emotions evoked by subliminal words. *Proceedings of the National Academy of Sciences of the United States of America* **102** (21): 7713-7717, 2005.

Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., Hatano, K., Ito, K., Fukuda, H., Schormann, T., and Zilles, K. Functional delineation of the human occipito-temporal areas related to face and scene processing. A PET study. *Brain : a journal of neurology* **123** (Pt 9): 1903-1912, 2000.

Nishitani, N. and Hari, R. Temporal dynamics of cortical representation for action. *Proceedings of the National Academy of Sciences of the United States of America* **97** (2): 913-918, 2000.

Ochsner, K.N., Ray, R.D., Cooper, J.C., Robertson, E.R., Chopra, S., Gabrieli, J.D., and Gross, J.J. For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. *NeuroImage* **23** (2): 483-499, 2004.

Pelphrey, K.A., Morris, J.P., and McCarthy, G. Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of cognitive neuroscience* **16** (10): 1706-1716, 2004.

Pelphrey, K.A., Singerman, J.D., Allison, T., and McCarthy, G. Brain activation evoked by perception of gaze shifts: the influence of context. *Neuropsychologia* **41** (2): 156-170, 2003.

Petrini, K., Pollick, F.E., Dahl, S., McAleer, P., McKay, L.S., Rocchesso, D., Waadeland, C.H., Love, S., Avanzini, F., and Puce, A. Action expertise reduces brain activity for audiovisual matching actions: an fMRI study with expert drummers. *NeuroImage* **56** (3): 1480-1492, 2011.

Pichon, S., de Gelder, B., and Grezes, J. Two different faces of threat. Comparing the neural systems for recognizing fear and anger in dynamic body expressions. *NeuroImage* **47** (4): 1873-1883, 2009.

Porges, E.C. and Decety, J. Violence as a source of pleasure or displeasure is associated with specific functional connectivity with the nucleus accumbens. *Frontiers in human neuroscience* **7**: 447, 2013.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. Premotor cortex and the recognition of motor actions. *Brain research. Cognitive brain research* **3** (2): 131-141, 1996.

Ruby, P. and Decety, J. Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nature neuroscience* **4** (5): 546-550, 2001.

Sawahata, Y., Khosla, R., Komine, K., Hiruma, N., Itou, J., Watanabe, S., Suzuki, Y., Hara, Y., and Issiki, N. Determining comprehension and quality of TV programs using eye-gaze tracking. *Pattern Recogn* **41** (5): 1610-1626, 2008.

Saxe, R. Why and how to study Theory of Mind with fMRI. *Brain research* **1079** (1): 57-65, 2006.

Saxe, R. and Kanwisher, N. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *NeuroImage* **19** (4): 1835-1842, 2003.

Saxe, R., Xiao, D.K., Kovacs, G., Perrett, D.I., and Kanwisher, N. A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* **42** (11): 1435-1446, 2004.

Schiffer, A.M., Ahlheim, C., Ulrichs, K., and Schubotz, R.I. Neural changes when actions change: adaptation of strong and weak expectations. *Human brain mapping* **34** (7): 1713-1727, 2013.

Schultz, R.T., Grelotti, D.J., Klin, A., Kleinman, J., Van der Gaag, C., Marois, R., and Skudlarski, P. The role of the fusiform face area in social cognition: implications for the pathobiology of autism. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **358** (1430): 415-427, 2003.

Seger, C.A. and Cincotta, C.M. Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cereb Cortex* **16** (11): 1546-1555, 2006.

Shah, A., Jhawar, S.S., and Goel, A. Analysis of the anatomy of the Papez circuit and adjoining limbic system by fiber dissection techniques. *J Clin Neurosci* **19** (2): 289-298, 2012.

Shane, M.S., Stevens, M.C., Harenski, C.L., and Kiehl, K.A. Double dissociation between perspective-taking and empathic-concern as predictors of hemodynamic response to another's mistakes. *Social cognitive and affective neuroscience* **4** (2): 111-118, 2009.

Smith, T.J. and Mital, P.K. Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of vision* **13** (8), 2013.

Soto, D., Humphreys, G.W., and Rotshtein, P. Dissociating the neural mechanisms of memory-based guidance of visual selection. *Proceedings of the National Academy of Sciences of the United States of America* **104** (43): 17186-17191, 2007.

Stadler, W., Schubotz, R.I., von Cramon, D.Y., Springer, A., Graf, M., and Prinz, W. Predicting and memorizing observed action: differential premotor cortex involvement. *Human brain mapping* **32** (5): 677-687, 2011.

Stainer, M.J., Scott-Brown, K.C., and Tatler, B.W. Looking for trouble: a description of oculomotor search strategies during live CCTV operation. *Frontiers in human neuroscience* **7**: 615, 2013.

Stedmon, A., Harris, S., Carse, A., Sharples, S., and Wilson, J. Tracking a suspicious person using CCTV: but what do we mean by suspicious behaviour? In P. Bust (ed) *Contemporary Ergonomics* 2008. Taylor & Francis: London., 2008.

Stern, C.E., Corkin, S., Gonzalez, R.G., Guimaraes, A.R., Baker, J.R., Jennings, P.J., Carr, C.A., Sugiura, R.M., Vedantham, V., and Rosen, B.R. The hippocampal formation participates in novel picture encoding: evidence from functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America* **93** (16): 8660-8665, 1996.

Strafella, A.P. and Paus, T. Modulation of cortical excitability during action observation: a transcranial magnetic stimulation study. *Neuroreport* **11** (10): 2289-2292, 2000.

Strenziok, M., Krueger, F., Deshpande, G., Lenroot, R.K., van der Meer, E., and Grafman, J. Fronto-parietal regulation of media violence exposure in adolescents: a multi-method study. *Social cognitive and affective neuroscience* **6** (5): 537-547, 2011.

Tomeo, E., Cesari, P., Aglioti, S.M., and Urgesi, C. Fooling the Kickers but not the Goalkeepers: Behavioral and Neurophysiological Correlates of Fake Action Detection in Soccer. *Cereb Cortex*, 2012.

Troscianko, T., Holmes, A., Stillman, J., Mirmehdi, M., Wright, D., and Wilson, A. What happens next? The predictability of natural behaviour viewed through CCTV cameras. *Perception* **33** (1): 87-101, 2004.

Turk-Browne, N.B., Yi, D.J., and Chun, M.M. Linking implicit and explicit memory: common encoding factors and shared representations. *Neuron* **49** (6): 917-927, 2006.

Urry, H.L., van Reekum, C.M., Johnstone, T., Kalin, N.H., Thurow, M.E., Schaefer, H.S., Jackson, C.A., Frye, C.J., Greischar, L.L., Alexander, A.L., and Davidson, R.J. Amygdala and ventromedial prefrontal cortex are inversely coupled during regulation of negative affect and predict the diurnal pattern of cortisol secretion among older adults. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **26** (16): 4415-4425, 2006.

Van Hoesen, G.W. The parahippocampal gyrus: New observations regarding its cortical connections in the monkey. *TINS* **5**: 345-350, 1982.

Van Overwalle, F. and Baetens, K. Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *NeuroImage* **48** (3): 564-584, 2009.

Vuilleumier, P., Armony, J.L., Driver, J., and Dolan, R.J. Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. *Neuron* **30** (3): 829-841, 2001.

Wager, T.D. and Nichols, T.E. Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *NeuroImage* **18** (2): 293-309, 2003.

Weinberger, S. Airport security: Intent to deceive? *Nature* **465** (7297): 412-415, 2010.

Welsh, B. and Farrington, D. Crime prevention effects of closed circuit television: a systematic review. *Home Office Research Study 252, London: HMSO.*, 2002.

Weniger, G., Boucsein, K., and Irle, E. Impaired associative memory in temporal lobe epilepsy subjects after lesions of hippocampus, parahippocampal gyrus, and amygdala. *Hippocampus* **14** (6): 785-796, 2004.

Zola-Morgan, S., Squire, L.R., Amaral, D.G., and Suzuki, W.A. Lesions of perirhinal and parahippocampal cortex that spare the amygdala and hippocampal formation produce severe memory impairment. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **9** (12): 4355-4370, 1989.

Figure captions

Fig. 1. Signal detection analysis applied to participants' responses. The first diagram on the left represents the number of hits (rightly judging a display as having a violent outcome) as a function of false alarms (wrongly judging a display as having a violent outcome). The blue triangles represent the responses for the 15 operators, and the red circles those for the 14 novices. The dashed line represents the 'random guess' or chance level of response, in the sense that points above this line indicate performance better than chance, and points below this worse than chance. The second and third diagrams from the left represent the sensitivity index (d') and bias (criterion) for the operators and novices respectively. Error bars = standard errors.

Fig. 2. *fMRI results for clip types*: Clusters of activation for which the difference between the brain responses to the four types of clips ('fight', 'confront', 'play' and 'nothing') varied across the two groups of participants (operators and novices). Sagittal and transverse slices show activation foci for the parahippocampal gyrus and superior frontal gyrus (SFG) respectively. The detected foci of activations are shown on the average anatomical map obtained from all participants. The activity in the right parahippocampal gyrus (uncus and posterior parahippocampal gyrus) is shown on the left-hand side, with the activity in the left SFG on the right-hand side. The details of the foci and analysis are reported at the bottom of the figure. The average contrast parameters (beta weights) and relative standard errors are shown either in separate histograms or different colors for operators (white) and novices (black). Different colors are also used to represent the different clip types: 'fight' = black; 'confront' = light blue; 'play' = white; 'nothing' = yellow. The correlation between CCTV experience and contrast parameters as well as between perspective taking and contrast parameters was obtained by means of linear regression analyses. To view the clusters of activation in sagittal, coronal and transverse orientations please refer to Figure 2S in the supplemental material.

Fig. 3. *fMRI results for judged intent*: Clusters of activation for which the difference between the brain responses to the judged intent (fight or no-fight) in the clips varied across the two groups of participants (operators and novices). The sagittal slice shows the uncus activation (notice that this is exactly the same area we found with the previous analysis). The detected focus of activation is shown on the average anatomical map obtained from all participants. The details of the foci and analysis are reported at the bottom of the figure. The average

contrast parameters (beta weights) and relative standard errors are shown in separate histograms for operators and novices and different colors for clips judged as ending in a fight (black) and clips judged as not ending in a fight (white). The correlation between CCTV experience and contrast parameters was obtained by means of linear regression analyses. To view the clusters of activation in sagittal, coronal and transverse orientations please refer to Figure 3S in the supplemental material.

Fig. 4. *fMRI results for salient events:* Clusters of activation for which the difference between the brain responses to the most salient event in the clips (i.e. the time and spatial location when the observers' eye gazes were most in agreement with each other) varied across the two groups of participants (operators and novices). Sagittal slices show activation foci for the inferior temporal gyrus (ITG, on the top left-hand side), superior occipital gyrus (SOG), right parahippocampal gyrus (on the top right-hand side), left parahippocampal gyrus (on the bottom left-hand side), and inferior occipital gyrus (IOG, on the bottom right-hand side). The detected foci of activations are shown on the average anatomical map obtained from all participants. The details of the foci and analysis are reported at the bottom of the figure. The average contrast parameters (beta weights) and relative standard errors are shown as different colors for operators (white) and novices (black). To view the clusters of activation in sagittal, coronal and transverse orientations please refer to Figure 4S in the supplemental material.

Fig. 5. *fMRI results for intersubject correlation analysis:* Whole-brain statistical comparisons of intersubject correlations for operators-to-operators with novices-to-novices (independent samples t-tests over individual subject mean correlation maps). Number of voxels and Brodmann area (BA) for the five brain regions showing higher correlation for operators than novices (in orange): right superior temporal gyrus (STG) = 309 voxels and BA 38; left ventral striatum = 833 voxels; left superior temporal gyrus (STG) = 1946 voxels and BA 38; left middle temporal gyrus (MTG) = 448 voxels and BA 39; left inferior parietal lobule (IPL) = 1339 voxels and BA 39. The values between parentheses represent the Talairach coordinates (x, y, z). To view the clusters of activation in sagittal, coronal and transverse orientations please refer to Figure 5S in the supplemental material. The brain regions showing higher correlation for novices than operators are shown in blue.

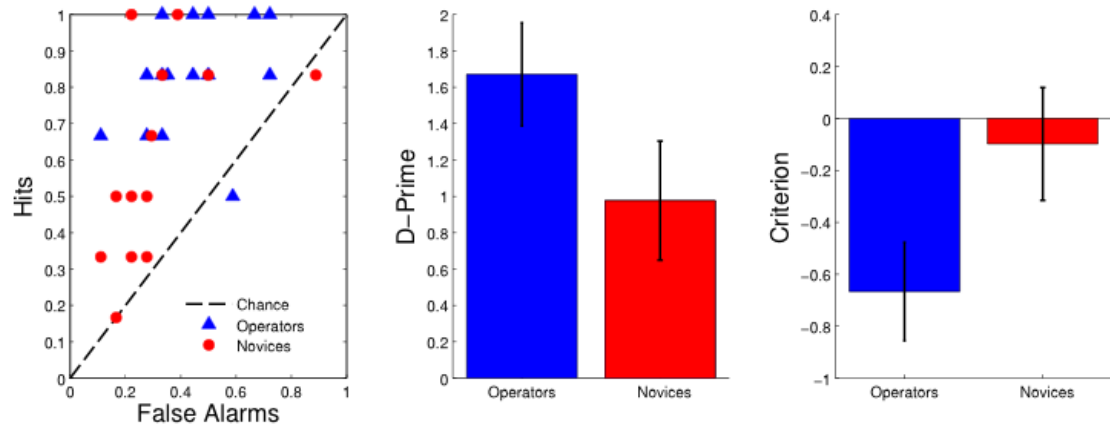
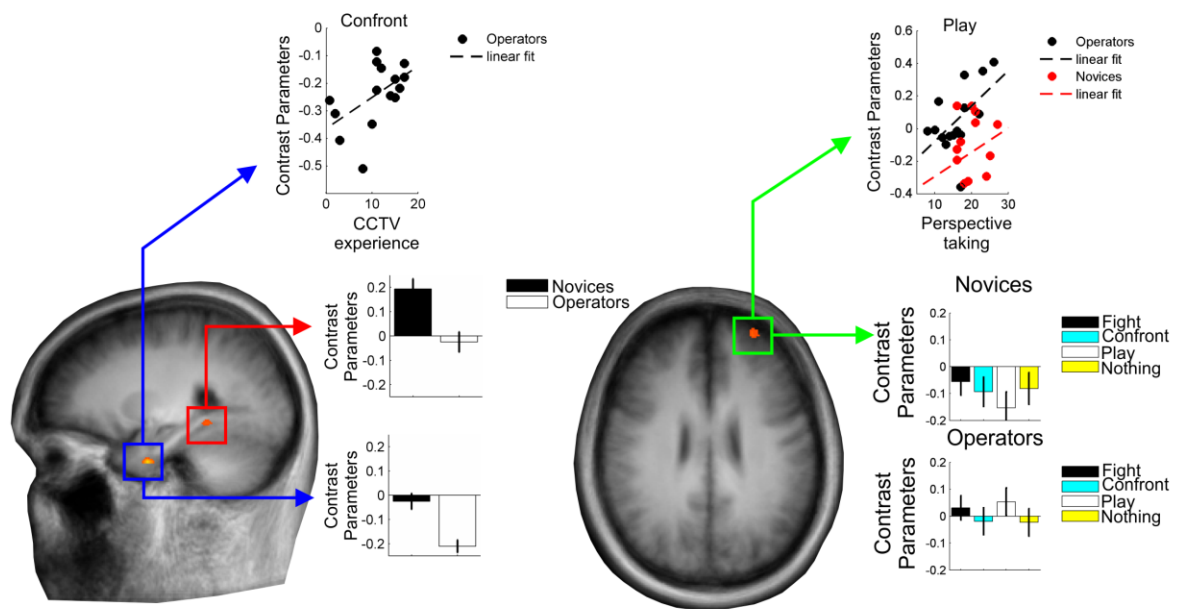


Figure 2



Anatomical region	Hemisphere	Talairach – coordinate (x,y,z)	Number of voxels	Effect size*		BA
				<i>F</i> (1, 28)	<i>P</i>	
<i>Experience</i>						
Parahippocampal gyrus	Right	25, -40, -6	104	15.04	0.0006	36
Uncus	Right	26, -2, -30	323	18.09	0.0003	36
<i>Experience x Clip Type</i>				<i>F</i> (3, 84)	<i>P</i>	
SFG	Left	-25, 53, 28	131	6.49	0.0005	9

*Effect size = average F value for all voxels in the ROI. Legend: BA – Brodmann's area; SFG = superior frontal gyrus.

Figure 3

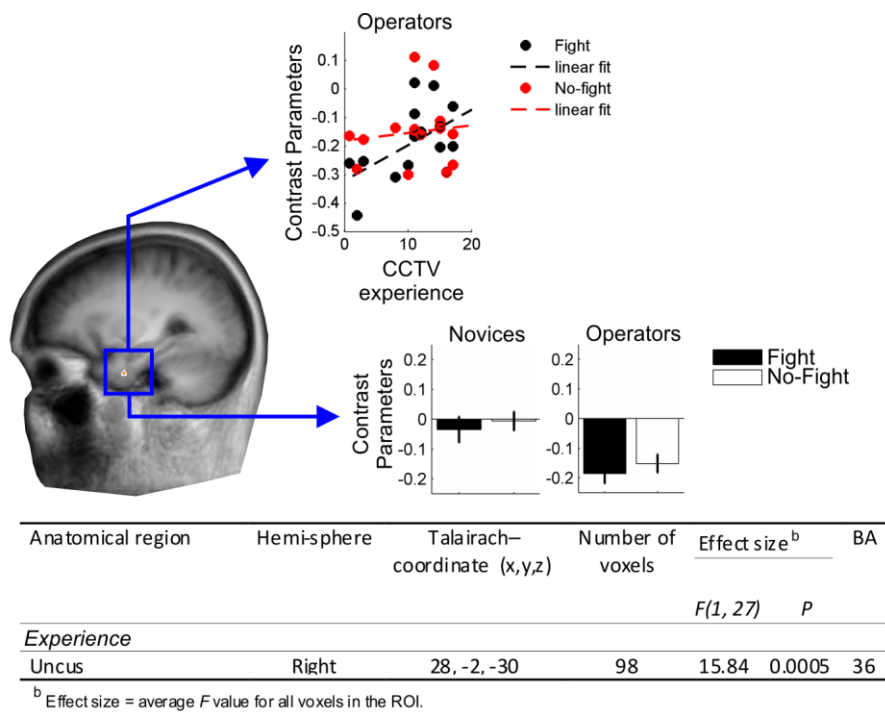


Figure 4

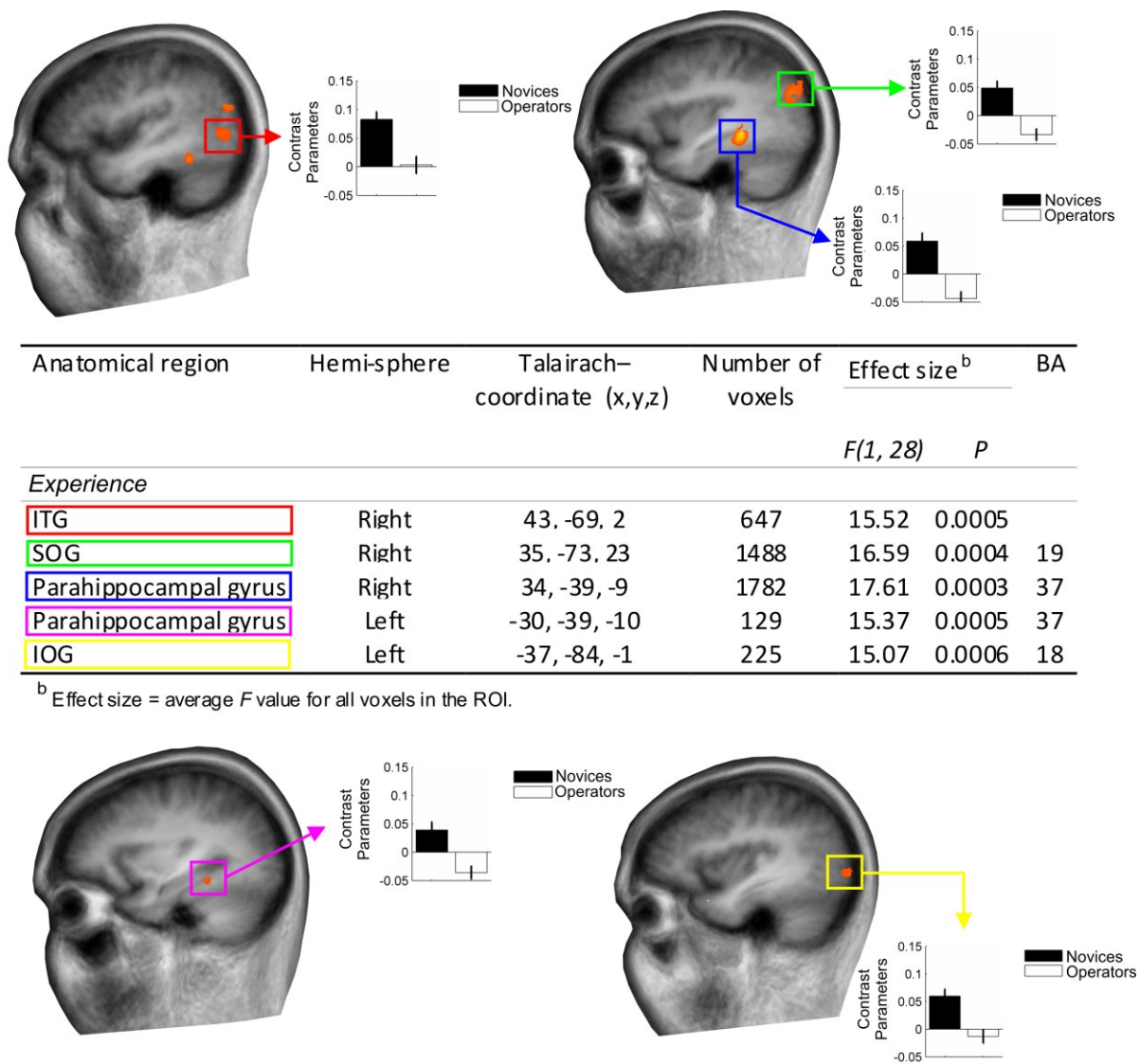


Figure 5

